Proceedings of the 37th IEEE
Conference on Decision & Control
Tampa, Florida USA • December 1998

Analytic Center and Maximum Likelihood Estimators

Er-Wei Bai[1], Minyue Fu[2], Roberto Tempo[3] and Yinyu Ye[4]

[1]Dept of ECE, Univ. of Iowa, [2]Dept of ECE, Univ. of Newcastle, [3]CENS-CNR, Italy, [4]Dept of Manag Scie, Univ of Iowa

Abstract

The purpose of this note is to show that the analytic center, originally proposed for identification in a bounded error setting, is a Maximum Likelihood Estimator for a class of probability distributions.

1. Introduction

Consider a discrete-time system

$$y = \Phi\theta + v, \tag{1.1}$$

with

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ \Phi = \begin{pmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_n^T \end{pmatrix} \text{ and } v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

where $y_i \in \mathbf{R}$ is the system output, $\phi_i \in \mathbf{R}^m$ the measurable regressor, $\theta \in \mathbf{R}^m$ the unknown parameter vector to be identified and $v_i \in \mathbf{R}$ the noise.

The purpose of parametric system identification is to find an estimate $\hat{\theta}$ of the unknown parameter vector θ from available input-output measurements y and Φ. Throughout the paper, the noise is assumed to be unknown but bounded by

$$|v_i| \leq \epsilon \tag{1.2}$$

for some known $\epsilon > 0$ and $i = 1, 2, \ldots, n$. Then, the membership set

$$\Omega^n = \bigcap_{i=1}^{n} \{\theta \in \mathbf{R}^m \ : \ -\epsilon \leq y_i - \phi_i^T\theta \leq \epsilon\} \tag{1.3}$$

is the set of all parameter estimates that are consistent with the equation (1.1), the input-output data and the assumed noise bound (1.2). Several papers studied specific estimates in the membership-set enjoying certain optimality properties. For example, a well-known estimate is the Chebyshev θ_c center of the set Ω^n

$$\theta_c = \arg \min_{\theta \in \Omega^n} \max_{\eta \in \Omega^n} \|\theta - \eta\|.$$

Another well-known estimate is the projection estimate θ_p, also denoted as the constrained least squares estimate,

$$\theta_p = \arg \min_{\hat{\theta} \in \Omega^n} \sum_{i=1}^{n} (y_i - \phi_i^T\hat{\theta})^2 = \arg \max_{\hat{\theta} \in \Omega^n} \sum_{i=1}^{n} (\epsilon^2 - (y_i - \phi_i^T\hat{\theta})^2).$$

In [1], [2] and [3] an analytic center approach was proposed for set membership identification; see also [4] for a definition of analytic center. In this approach, a specific estimator, the analytic center that is within the membership-set minimizing the logarithmic-average output error or, equivalently, maximizing the complementary logarithmic-average output error

$$\theta_a = \arg \max_{\hat{\theta} \in \Omega^n} \Pi_{i=1}^{n} (\epsilon^2 - (y_i - \phi_i^T\hat{\theta})^2)$$

$$= \arg \max_{\hat{\theta} \in \Omega^n} \sum_{i=1}^{n} \ln(\epsilon^2 - (y_i - \phi_i^T\hat{\theta})^2)$$

is proposed. Moreover, it was shown in [1] that the computation of a sequence of analytic centers at times $i = 1, 2, \ldots, n$ can be made sequential. In particular, the total cost, in terms of Newton iterations, to compute a sequence of analytic centers up to n, is linear in n. This is a very attractive feature for on-line identification.

2. Analytic Center and Maximum Likelihood Estimators

The analytic center minimizes the logarithmic-average output error in a deterministic sense. In this section, we show that the analytic center is also a Maximum Likelihood Estimator (see [5] for definition) for a class of probability density functions. To this end, we consider a class of probability density functions $q(v_i)$ with support bounded in the interval $[-\epsilon, \epsilon]$

$$q(v_i) = \begin{cases} b(r)(\epsilon^2 - v_i^2)^r & |v_i| \leq \epsilon \\ 0 & |v_i| > \epsilon \end{cases} \tag{2.1}$$

for some real $r \geq 0$, where $b(r)$ is a normalizing constant. Then, we have

Theorem 2.1 *Consider the system (1.1). Suppose that the noise v_i is a sequence of identically independently distributed random variables independent of the regressor Φ with the density function $q(v_i)$ belonging to the set (2.1) for some $r \geq 0$. Then, the analytic center is a Maximum Likelihood Estimator. More precisely, let $q(y_1, y_2, \ldots, y_n|\theta)$ be the joint probability density function of the random variables $y_1, y_2, \ldots, y_n$ for given θ. Then,*

$$\theta_a = \arg \max_{\hat{\theta} \in \mathbf{R}^m} q(y_1, y_2, \ldots, y_n|\hat{\theta}).$$

We now state some remarks about the set (2.1) and the theorem:

- The set (2.1) includes parabolic density functions for which $r = 1$. We also notice that several standard density functions can be represented or approximated by members of the above class. For example, in the case $r = 0$, $q(v_i)$ coincides with the uniform distribution in the interval $[-\epsilon, \epsilon]$. The triangle distribution can be approximated by members of the set (2.1). In addition, truncated Gaussian distributions with zero mean can also be approximated. For example, let v_i be a random variable with zero mean Gaussian distribution truncated in the interval $|v_i| \leq \alpha\sigma$ for some $\alpha > 0$. Then, its distribution is given by

$$q(v_i) = \begin{cases} \frac{b_1}{\sqrt{2\pi}\sigma}\left(e^{-\frac{v_i^2}{2\sigma^2}} - e^{\frac{-(\alpha\sigma)^2}{2\sigma^2}}\right) & |v_i| \leq \alpha\sigma \\ 0 & |v_i| > \alpha\sigma \end{cases}$$

$$= \begin{cases} \frac{b_1 e^{\frac{-\alpha^2}{2}}}{\sqrt{2\pi}\sigma}\left(e^{\frac{\alpha^2}{2} - \frac{v_i^2}{2\sigma^2}} - 1\right) & |v_i| \leq \alpha\sigma \\ 0 & |v_i| > \alpha\sigma \end{cases}$$

where b_1 is a normalizing constant. Since $e^{\frac{\alpha^2}{2} - \frac{v_i^2}{2\sigma}} - 1 \approx \frac{\alpha^2}{2} - \frac{v_i^2}{2\sigma^2}$, the first order approximation of $q(v_i)$ becomes

$$q(v_i) \approx \begin{cases} b_2\left(\frac{\alpha^2}{2} - \frac{v_i^2}{2\sigma^2}\right) & |v_i| \leq \alpha\sigma \\ 0 & |v_i| > \alpha\sigma \end{cases}$$

$$= \begin{cases} b(r)(\epsilon^2 - v_i^2)^r & |v_i| \leq \epsilon \\ 0 & |v_i| > \epsilon \end{cases}$$

where $b(r) = b_2/(2\sigma^2)$, $\epsilon = \alpha\sigma$ and $r = 1$. Therefore, the first order approximation of the truncated Gaussian is in the set (2.1) for some $b(r), \epsilon$ and r. Clearly, if the first order approximation is not sufficient, we can easily construct a second order or, more generally, a higher order approximation. It can be shown that higher order approximations of the zero mean truncated Gaussian density functions are also members of the set (2.1). In Figure 1, we show a uniform, a triangle and two zero mean truncated Gaussian density functions and their approximations by some elements in the set (2.1). The top picture shows the uniform distribution in the interval $[-1, 1]$ (dash-dot line) and the distribution (solid line) of the element in the set (2.1) with $r = 0$. The second picture shows the triangle distribution in the interval $[-1, 1]$ (dash-dot line) and its approximation (solid line) by an element in the set (2.1) with $r = 2.6$. The third one is the zero mean truncated Gaussian (dash-dot line) of $\alpha = 1$ and $\sigma = 1$ and its first order approximation (solid line) by an element in the set (2.1) with $r = 1$, $\epsilon = \alpha\sigma = 1$ and $b(1) = 0.7501$. The bottom one shows the zero mean truncated Gaussian (dash-dot line) of $\alpha = 3$ and $\sigma = 1/3$ and its higher order approximation (solid line) by an element in the set (2.1) with $r = 4$ and $b(4) = 1.2305$.

- If the noise v_i is a sequence of identically independently distributed zero mean Gaussian, the Least Squares estimate is obviously the Maximum Likelihood Estimator. If the noise v_i is a sequence of identically independently distributed zero mean truncated Gaussian, however, the Least Squares estimate is no longer the Maximum Likelihood Estimator. On the other hand, as discussed above, the analytic center approximately coincides with the Maximum Likelihood Estimator when the noise v_i is a sequence of identically independently distributed truncated Gaussian distributions with zero mean.

References

[1] Bai, E.-W., Y. Ye and R. Tempo (1997). "Bounded Error Parameter Estimation: A Sequential Analytic Center Approach," *Proceedings of the IEEE Conference on Decision and Control*, San Diego; also to appear in *IEEE Transactions on Automatic Control.*

[2] Bai, E.-W., M. Fu, R. Tempo and Y. Ye (1998). "Analytic Center Approach to Parameter Estimation: Convergence Analysis," *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, Seattle.

[3] Bai, E.-W., R. Tempo and Y. Ye (1998). "Recent Results on the Analytic Center Approach for Bounded Error Parameter Estimation," to appear in Learning Control and Hybrid Systems, Y. Yamamoto and S. Hara Editors, Springer-Verlag, London 1998.

[4] Ye Y. (1997). *Interior-Point Algorithm: Theory and Analysis*, Wiley, New York.

[5] Ljung, L. (1987). *System Identification: Theory for the Users*, Prentice-Hall, Englewood Cliffs, NJ.

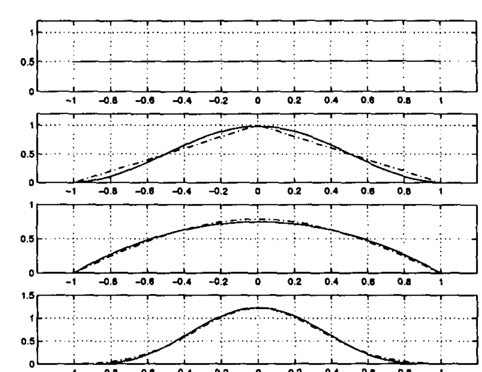

Figure 1